

A Pilot Year in Review

WHAT HAVE WE LEARNED ABOUT
THROUGH-YEAR ASSESSMENTS?

About this publication

A Pilot Year in Review: What have we learned about through-year assessments?

This publication interrogates key takeaways from through-year assessment pilots administered during the 2022-2023 school year. We explore key design decisions, enabling conditions and implications for future research and practice. This publication is part of a series published through Education First's Through-year Curriculum-Connected Assessment Grant Program.

Authors



Dave Powell
Senior Consultant
Education First



Emma Fortier
Associate
Education First

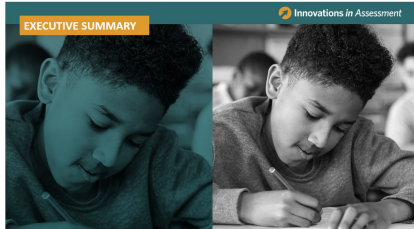


Senna Lamba
Associate
Education First



Khaled Ismail
Principal
Education First

Directory



Executive Summary



Why Through-Year?



Lessons Learned



Full Report

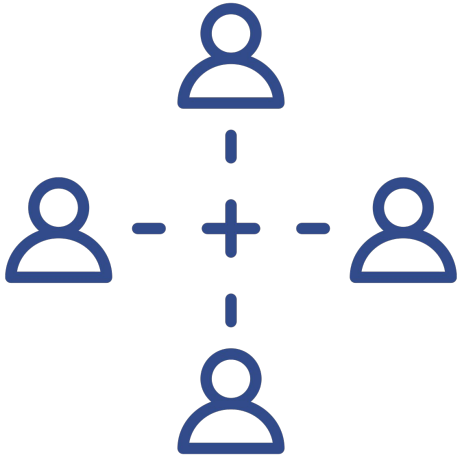


Exploring the Models



[Return to the landing page](#)

Table of Contents



01. Exploring the Models

1a. Through-Year: CenterPoint's Curriculum-Aligned Interim Assessments

1b. CrawFish Model

1c. MasteryGuide Assessments

02. Appendix

Click on these titles to navigate to other sections within this presentation.



*While state summative assessments serve an important role in our education system, **they have the potential to improve through various innovations***

Education First believes students, educators, families and state leaders need more equitable, focused and relevant assessments that strengthen the connection between assessment and instruction and better align what is tested with what is taught

Since 2019, we've led grant and coaching programs to advance innovations in assessment, reporting and accountability



Grants have supported a range of innovations, including **through-year assessments, computer-adaptive assessments, comprehensive graduate portfolios, whole-child measures and equity indicators.**

The curriculum-connected through-year assessment grant program has focused on:

Incentivizing R&D to support states to focus on curriculum-connected through-year models

Building connections and buy-in among federal advocates and policymakers for change, including supporting CGSA grant writing

Facilitating a community of practice among grantees, innovative states and developers pursuing through-year assessment

Sharing our learnings and thought leadership with the field



In this report, we use the following definitions that build on our prior thinking and the work of others in this field:

Through-Year Assessment Models

Through-year assessment models administer multiple tests throughout the school year as part of an assessment system designed to produce a single summative score meeting federal and state accountability requirements. Through-year assessment models are also referred to as “through-course” by some states.

Curriculum-Agnostic Approach

Through-year assessment models that **test the entire content domain (or grade-level standards) throughout the year at each testing administration**, and do not try to align content tested to curriculum.

Curriculum-Aligned Approach*

Through-year assessment models that directly **draw on the content found in specific curriculum**. This model is also referred to as “curriculum-specific” or “curriculum-embedded.”

Curriculum-Relevant Approach

Through-year assessment models that can be **flexibly aligned with multiple curricula, a scope and sequence or pacing of content**. This approach is also referred to as “scope and sequence aligned”, “instructionally relevant” or “instructionally aligned.”

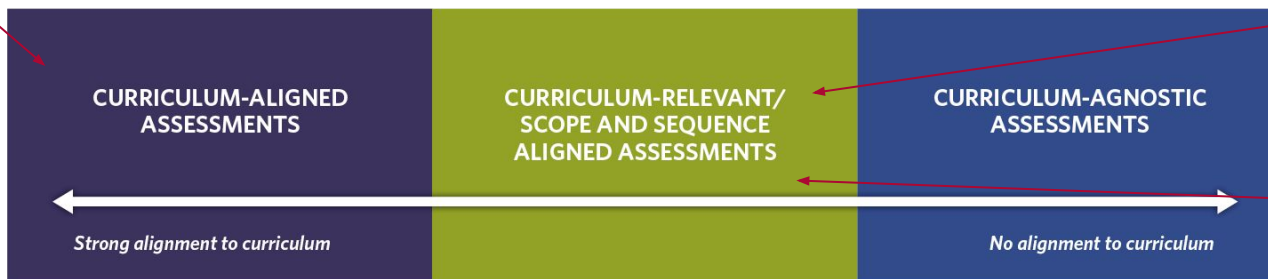
*In previous publications, we referred to this as curriculum-specific embedded

EXPLORING THE MODELS

HOW HAVE STATES AND ASSESSMENT DEVELOPERS APPROACHED THE CHALLENGE OF ALIGNING THROUGH-YEAR ASSESSMENTS TO CURRICULUM AND/OR SCOPE AND SEQUENCES?



For the past year, we funded three assessment developers to pursue curriculum-connected, through-year assessments



Adapted from Dadey and Badrinarayan (2022)

In this section, we provide a deep dive into these three models and cover the following questions:

1



What is the developer trying to accomplish?

2



How is the model accomplishing this?

3



What did the developer learn so far and how is that influencing their plans moving forward?

Through-Year:
CenterPoint's
Curriculum-Aligned
Interim
Assessments



CenterPoint's Curriculum-Aligned Interim Assessments are based on the premise that students can effectively demonstrate their knowledge through assessments that are closely tied to instruction

Overview

CenterPoint is conducting research on the **feasibility of a curriculum-aligned through-year model aligned to Illustrative Math (IM)**

Model

Curriculum-Aligned

Goals

CenterPoint aims to answer the following research questions in this study:

- Do the CenterPoint IM Interim assessments measure similar learning outcomes as summative assessments?
- Do the CenterPoint IM Interim assessments predict students' performance on summative assessments?

Partners

CenterPoint collaborated with two K-12 urban districts located in distinct regions of the U.S. collectively serving around 160,000 students. Both districts serve diverse student populations.

What problem is this model solving?

CenterPoint believes that students can effectively demonstrate their knowledge through assessments that are closely tied to curriculum and instruction

CenterPoint's study addresses the challenge of potential misalignment between summative assessments and instruction in the classroom



The Challenge

There is a **potential misalignment between summative assessments and a curriculum's scope and sequence**, instruction and the measurement of those learnings by interim assessments. This exemplifies the **tension between when and how often students show mastery**. Educators try hard to navigate the tension by supporting and valuing each of the assessments in their suite. Yet, the substantial emphasis on the single end-of-year measure remains a hurdle.



The Solution

By analyzing the CenterPoint Illustrative Mathematics Interim assessments alongside summative assessments, CenterPoint observed commonalities in intended measures and outcomes, spanning standards, domains and results. This enables them to assess predictability for the summatives and **investigate the potential of the interim assessments to become the primary indicator of student learning**. Assisting educators in understanding these interrelationships, coupled with guidance in navigating this understanding, will help educators effectively assist their students.

CenterPoint aims to create a curriculum-aligned through-year assessment tied to the Illustrative Mathematics curriculum



Model Overview

- CenterPoint chose the Illustrative Mathematics (IM) curriculum for math assessments because it is comprehensive and rated green on Ed Reports. CenterPoint's IM-aligned interim assessments have received certification from the IM organization.
- The assessments are designed to be administered three times a year (fall, winter, spring) to provide a comprehensive picture of students' progress within the curriculum.
- The current study has thus far focused its efforts on the middle school grades.
- The ultimate goal is to broaden the scope of the study and its reach to include multiple states to include CenterPoint's interims in kindergarten through high school These assessments include:
 - 3-4 interims per grade level
 - Each interim takes approximately 50 minutes
 - Each interim includes a variety of item types and cognitive complexity levels



CenterPoint conducted a study to determine the feasibility of generating a summative score using their IM-aligned interims



Study Design

This study aimed to determine the concurrent validity between the two assessments (to what degree they measure the same outcome).

- They analyzed results from two district's CenterPoint IM interim assessments against those districts' summative results from 2021-2022 school year using an IRT (item response theory) model.
- This included placing the results from both the summative and IM assessments on a common scale to determine the degree to which the results are correlated.
- The objective was to analyze the “thetas,” or student abilities, in their performance for both the interim and summative assessments

Findings

The initial findings demonstrated varying levels of correlation between CenterPoint's Illustrative Mathematics Curriculum-Aligned Interim Assessments and the summative assessments of the two urban districts. The analysis “**offer initial evidence that the CenterPoint IM interim assessments have the potential to yield similar summative outcomes.**”

Dr. Hong Jiao, Professor at the University of Maryland, College Park collaborated with CenterPoint on the psychometric analyses of this study



As CenterPoint moves forward, they are considering how to better meet students' and educators' needs in their design choices

In response to feedback from people most proximate to student learning, CenterPoint is considering some design changes in their assessments. Some of the potential design changes include:



Providing students **more opportunities to demonstrate their modeling and reasoning skills**, further assessing the breadth of the curriculum

Offering more **choice in the types of topics or context related to mathematics**

Building in **additional opportunities for culture and community**

Developing **additional professional learning opportunities to support educators and students**

Moving forward, CenterPoint has plans to increase the scope of their study and further explore growth measures and scale scores



CenterPoint's Future Research Plans are to:

Conduct a similar study involving additional testing grades for mathematics

Extend this study to other CenterPoint Curriculum-Aligned Interim Assessments

Explore the possibility of using standard setting to create a scale score similar to summative assessments, allowing for a **more detailed comparison between interim and summative assessments**

Investigate the potential for **measuring growth over time**

Consider the practicality of linking items between interim and summative assessments to **establish a meaningful connection between content and skills**





Louisiana hopes assessing students on content related to what they were taught will promote deeper engagement with texts and reduce disparities generated from students' varying levels of background knowledge

Overview

In partnership with Louisiana, NWEA created the CrawFish model so that the state's through-year assessments for ELA could **work for two HQIM: Guidebooks and Wit & Wisdom**

Model

Curriculum-Relevant

Goals

Louisiana and NWEA aims to accomplish the following goals through their work:

1. Provide the same level of, or better, information to educators (instructionally-oriented and practical, actionable information)
2. Incentivize deep engagement in the material and texts throughout the year

Partners

Louisiana Department of Education (LDOE): 23 districts; John Hopkins University, Center for Assessment, Odell Education

What problem is this model solving?

Create a curriculum-relevant, through-year assessment that aligns to all of the possible curricula within the state and provides useful information and results tied to curricula.



The CrawFish Model seeks to find a balance between curriculum-aligned and curriculum-agnostic through-year assessments in order to scale the solution statewide

Curriculum-Aligned

Louisiana created **curriculum-aligned** ELA through-year assessments for Guidebooks and Wit & Wisdom



The Challenge

Curriculum adoption varies across the state, so **curriculum-aligned models can not be used statewide**. It would be too expensive and impractical to develop curriculum-aligned models for all high-quality ELA curriculum.

- Many school systems adopt **more than one high-quality curriculum**
- Student movement between districts during the year can mean exposure to multiple curriculum.



Envisioned Solution

Develop an **ELA curriculum-relevant assessment** by drawing on common domains, topics, and texts found across curricula that could be used with any ELA curriculum—allowing Louisiana to **scale statewide** while **still incentivizing deep engagement with texts and reducing reliance on background knowledge**.

- Will complement or replace current curriculum-aligned assessments.

Louisiana planned to use “hot reads” of common texts across curriculum and “warm reads” of texts on related topics to achieve their goals



Hot reads allow students to engage with texts on the assessment which they have already studied in class, reducing the disparities generated by students' varying levels of background knowledge

Hot reads: texts students have read in class

Warm reads allow students to apply the common background knowledge on a topic generated from hot reads to a new text on the same topic

Warm reads: texts students have not read in class but are topically related to the information and knowledge they encountered in class



However, a curriculum analysis revealed that there were few common texts and topics across the five most commonly used curricula in Louisiana from which to build the CrawFish Model

From the curriculum analysis...

- **No texts were shared across all five curricula.**
 - While some texts were shared between two or three curricula at Grades 3 and 4, there was no overlap in Grade 5 for all five curricula.
- **Some topics were shared across the curricula**, such as marine animals and emotions, but overwhelmingly there were minimal shared topics.
- **Very little was shared across the curricula at sub-topic levels.**

This analysis was conducted by Dr. David Steiner and his team at the Institute for Education Policy of John Hopkins University.



As they couldn't use common texts across curricula, NWEA created two prototypes to find a statewide assessment model that honors background knowledge in a different way

A solution needs to:

- Align to all of the possible curricula within the state
- Provide useful information and results

CrawFish Funnel

The assessment builds background knowledge *based on shared themes across curricula*

Build Together Knowledge Model

The assessment builds background knowledge *within the test and draws on texts and learning from other subjects, such as social studies and science*



Louisiana and NWEA chose the CrawFish Funnel because educators supported how it retains the commitment to using background knowledge established in ELA classrooms

A solution needs to:

- Align to all of the possible curricula within the state
- Provide useful information and results



CrawFish Funnel

- The assessment uses background knowledge from different texts found in individual curriculum around a shared theme
- Educator feedback raised challenges with connecting an ELA assessment to background knowledge built in other subjects, as the Build Together Knowledge Model proposed.



The CrawFish Funnel honors background knowledge while connecting to different curriculum

1

**Unique, unit-based
“Knowledge” sections**

Hot read: Students see different texts here based on the curriculum they are using (e.g. from Guidebooks). All texts are thematically related.

Example: A text about the Civil Rights movement that students have already seen

2

A common thematically related warm read text and item set in the “Application” section

Warm read: All students see the same text here. It is thematically related to the text from section #1.

Example: A new text on the Civil Rights movement that students have not seen before

3

A common writing prompt in the “Synthesis” section

All students see the same prompt here. It is thematically related to the texts from sections #1 and #2.

Example: A question is posed about the Civil Rights movement

The scoring design will match what is currently used in the Innovative Assessment Program:
Data is pooled across units and scale scores are estimated using an IRT model



Despite roadblocks, Louisiana and NWEA prototyped one administration of the CrawFish model with almost 400 students as part of its broader ELA through-year pilots



Table 2: Sampling Design

Administration	Total Actual Count	Form A #	Form A %	Form B #	Form B %	Form C #	Form C %
WW	1,413	608	43	608	43	198	14
Guidebooks	2,422	1,114	46	1,114	46	194	8

“Form C” represents the CrawFish model prototype



Louisiana plans to administer a full-year pilot during the 23-24 school year and expand the curricula included in the CrawFish Funnel

In the 2023-2024 school year, the CrawFish Funnel pilot will be expanded to:

- A full-year pilot with three windows of administration
- Continue with Grade 5
- Connect to additional HQIM

The goals of the expanded pilot are to:

Recruit additional schools and additional curricula for the CrawFish funnel

Continue to refine report prototypes

Continue understanding how assessment data can be used to inform instruction

**MasteryGuide
Assessments**





New Meridian's flexibly administered testlets aim to provide actionable data to inform instruction throughout the year

Overview

New Meridian has created an assessment system of **short, modular “testlets”** in ELA and math **designed to be flexibly administered** across several administrations and **align with local scope and sequences**

Model

Scope and sequence aligned

Goals

- To create a ***coherent, continuous and useful assessment to better meet the goals of assessment for learning*** by administering frequent, mini-assessments aligned to local curriculum that provide **actionable data** to inform instruction throughout the year
- To replace traditional EOY statewide assessments and interims

Partners

Louisiana Department of Education (LDOE)

- Grades 5 & 7, math

Montana's Office of Public Instruction (OPI)*

- Grades 5 & 7, math & ELA

What problem is this model solving?

New Meridian's user research found the following problems with the current assessment system:

- Traditional statewide assessments return data too late to provide instructional value
- Interim assessments don't align to content taught and pull away from a coherent curriculum plan



New Meridian’s math testlets are organized by standard clusters and designed to align to local scope and sequence

Model Features

- The math model includes **12 individual testlets aligned to instructionally coherent clusters of standards**. The current design for the 2023 - 2024 year includes 14 testlets.
- Each testlets primarily assesses a single content strand. New Meridian designed items to assess different levels of cognitive depth within each strand.
- The math testlets are designed to be **spread across multiple administration windows**, and the number of testlets administered per window depends on districts’ local scope and sequence.

Initial Pilot

The math testlets assess content that could be taught in a variety of different sequences. **In this prototype, the order of administration was predetermined due to technical and logistical issues.**

This pilot only had machine-scored test items.



Future Plans

In future piloting, **educators will be able to choose from a bank of testlets and administer them in a schedule that works for their instructional timing.**

Future piloting will include **at least one constructed response math performance task** in a fifth administration, and the final design will include several performance task-based testlets.



The math testlets are designed to align to high-quality curricula and can be administered flexibly

- The math testlets each assess a single content strand and assess content that could *“reasonably be taught in a variety of different sequences.”* (New Meridian)
- Districts and schools will be able to choose the order of testlet administration to align with the order of content in their scope and sequence

Example: Aligning with Eureka

Module 1 of the Eureka 5th grade math scope and sequence focuses on place value and decimal fractions. Schools following this scope and sequence could administer the **Place Values- Power of 10** and **Place Values- Represent, Compare & Round** testlets as their first test administration at the end of this module.



Future Plans

The first unit of the Illustrative Math 5th grade scope and sequence focuses on Finding Volume. Schools following this scope and sequence could administer the **Units of Measurement** and **2 Dimensional Shapes** testlets as their first administration at the end of this module.



New Meridian's ELA testlets feature four administrations with increasingly complex text



Model Features

- ELA testlets assess **standards-based reading comprehension skills with increasingly complex passages and tasks.**
- In this pilot, there were **four sets of paired testlets given in four designated administration windows** - each testlet includes 1-2 passages
- The **testlets increase in textual complexity and in the skills assessed as the year progresses.**

Initial Pilot

In this pilot, all ELA test items were machine-scored.

New Meridian worked with the states to identify culturally relevant passages. In Montana, New Meridian included texts by Indigenous and rural authors.







Future Plans

Future piloting will include three sets of paired testlets administered in at least three separate windows, plus a writing-task based testlet

Future piloting will include **the development of multiple testlets with different textual complexities** for each administration to provide more flexibility for educators.



The ELA administrations increase in complexity of text and analysis throughout the year, following a developmental cognitive theory of acquisition of reading skills¹

Administration 1	Administration 2	Administration 3*	Administration 4
<ul style="list-style-type: none">▪ Includes readily accessible texts with explicit ideas▪ Evaluating single text elements 	<ul style="list-style-type: none">▪ Includes moderately complex texts that require light inferences▪ Some synthesis of texts 	<ul style="list-style-type: none">▪ Repeat of passages in Administration 1▪ Purpose is to measure progress 	<ul style="list-style-type: none">▪ Includes moderately to highly complex texts and items that require inferencing▪ Synthesizing texts 

*This describes the design for the small-scale pilot in the 2022-2023 school year. In the next pilot phase, the testlet in administration 3 will be a writing-task testlet



New Meridian is using this pilot database to inform the initial development of a summative scoring model

The model would generate a **mini-scale score based on a student's performance on an individual testlet**, and then **aggregate all the mini-scale scores into an overall summative score.**



The model will **weigh scores** based on the timing of the testlets' administration and their content.

New Meridian plans to develop summative and predictive scoring models to help Montana's OPI incorporate testlets into the state accountability system by the 2024-2025 school year.



New Meridian learned that users want expanded reports and actionable data to inform instruction

Concerns in initial pilot



Changes moving forward

User Experience



Users were generally positive about the functionality and usability of the platform. **Montana educators liked the student reports they received but requested expansions to the reporting system** including more details about the testlets and a classroom view of the data. States and district administrators requested reports at the school, district and state levels.



New Meridian is taking this feedback into account for the expanded pilot next year.

The Importance of Actionable Data



In this pilot, the testlets' order was predetermined and did not match some schools' scope and sequences. As a result, **students may have taken a math testlet assessing material they had not yet learned. This reinforced the importance of reporting actionable data in a timely manner**, as teachers do not want to waste time on assessments that provide information they cannot use.



The ability to tailor the administration of testlets to the local scope and sequence in a school system is important. This has influenced New Meridian's plans going forward, as we will describe in a following slide.



Inconsistency around test timing and the difficulty of test items had led New Meridian to make revisions

Concerns in initial pilot

Changes moving forward

Testlet Length



The amount of time estimated for a student to start and finish a test doesn't include all the administration time that is involved (e.g. logging in). **The testlets took longer than expected for some and the messaging around timing was inconsistent.** While New Meridian estimated that math testlets would take 10-15 minutes, OPI messaged to schools that it would take no more than 10 minutes, which led to inaccurate time expectations on teachers' end.

New Meridian is working to rapidly revise forms and plans on generating a timing estimate that includes non-assessment time.

Test Item Difficulty



Diagnostic analysis following each administration showed that most students were placed into certain skills profiles: **either showing they mastered every skill or no skills.** As a result, New Meridian recalibrated the assessment after the pilot closed, with more items in ELA and math. In ELA, New Meridian also adjusted the difficulty of the texts as student performance was closely tied to passage complexity. In math, New Meridian did a deep dive review of current test items to adjust the difficulty of items across testlets.

The math content team is updating guidelines for item writers to decrease the number of difficult items while still including enough of a range of difficulty to provide actionable feedback.



To address local scope and sequence, New Meridian plans to allow districts to select the order in which they administer testlets

In the next pilot, New Meridian plans to include a **configuration tool** that will allow schools/districts to schedule testlets to fit their local scope and sequence. **With this tool, schools and districts could choose the order of testlet administration.**

- The tool will **allow schools to enter the curriculum they use and would make a recommendation for the order of testlet administration**
 - The curriculum entered would need to be an identifiable enacted curriculum
- New Meridian is currently collecting curriculum data
- Focus groups with math and ELA educators to gather initial feedback on the prototype configurator tool helped determine what information educators want to see when scheduling testlets and provided insight into the reports educators need to see





Moving forward, New Meridian has plans for an expanded pilot in both Montana and Louisiana

In the 2023-2024 school year, New Meridian plans to expand the MasteryGuide Assessment pilot to



- Grades 3, 5, 6 and 7
- Math



- Grades 3 - 8
- Math & ELA

The goals of the expanded pilot are to:

Continue to develop the configurator tool to help districts align the ELA and math testlets to their unique scope and sequence

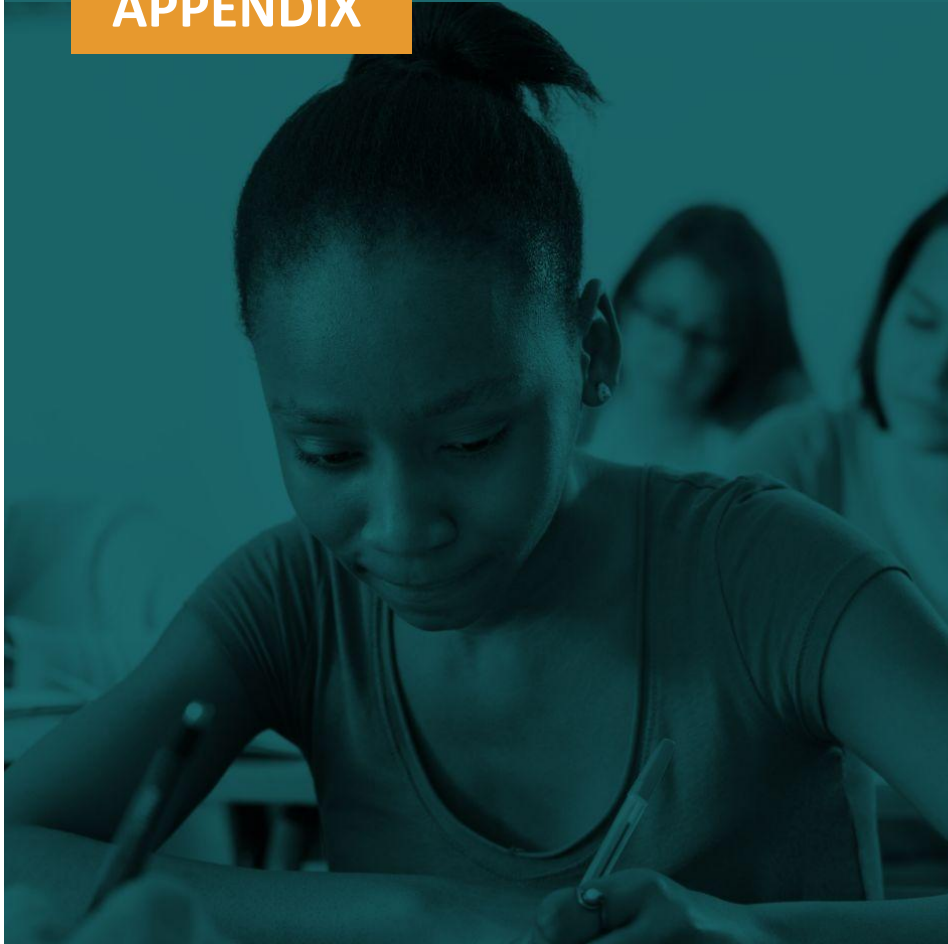
Improve usability based on user experience data

Continue to compare MasteryGuide's results with current summative assessments

Create additional items (in grades 3-8) to prepare for an operational field test in the 2024-2025 school year

Continue to develop expanded reporting tools

APPENDIX



Sources

Dadey, N., and Badrinarayan, A. (2022, April 21). In Search of the “Just Right” Connection Between Curriculum and Assessment: Considering Options Between Curriculum-Specific and Curriculum-Agnostic State Assessment. Center for Assessment.

<https://www.nciea.org/blog/in-search-of-the-just-right-connection-between-curriculum-and-assessment/>

Dadey, N., & Gong, B. (2023). An introduction to considerations for through-year assessment programs: purposes, design, development, evaluation [Research report]. Retrieved from Smarter Balanced website:

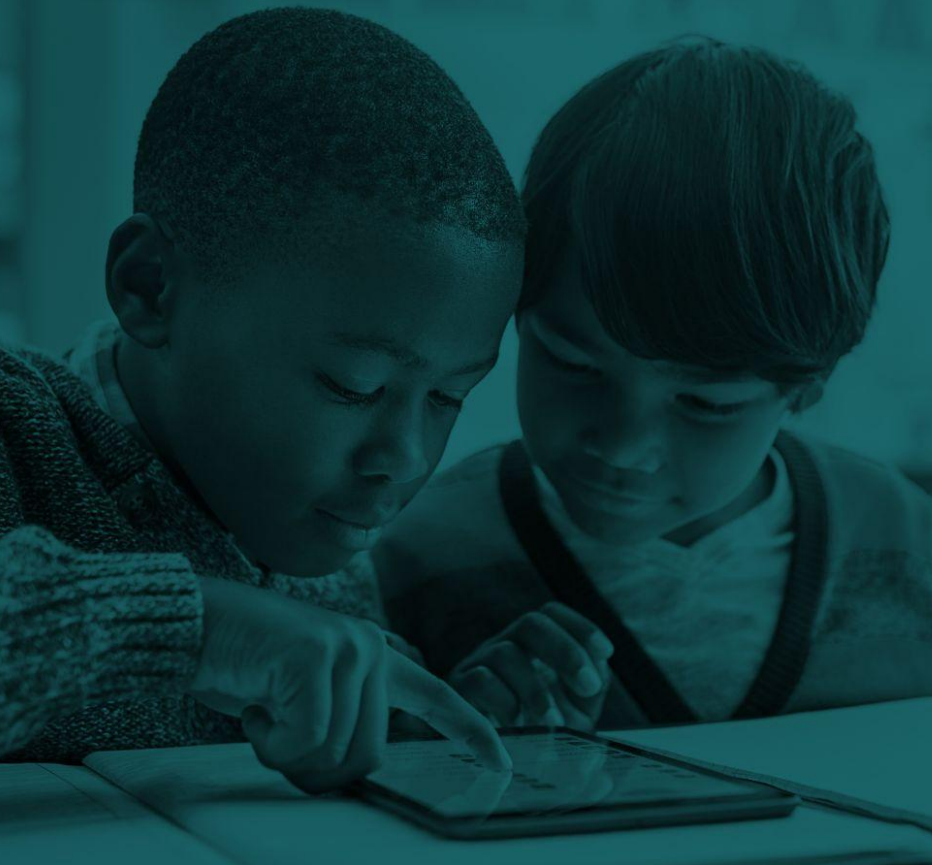
<https://portal.smarterbalanced.org/library/en/2023-sb-consideration-of-technical-issues.pdf>

Education First. (2022). What are through-year assessments?: Exploring multiple approaches to through-year design. Retrieved from

<https://www.education-first.com/wp-content/uploads/2023/01/What-are-Through-year-Assessments-1.pdf>

Marion, S. (2021, September 29). Trying to Serve Multiple Uses with Through Year Assessments. Center for Assessment.

THANK YOU!



Want to learn more?



Executive Summary



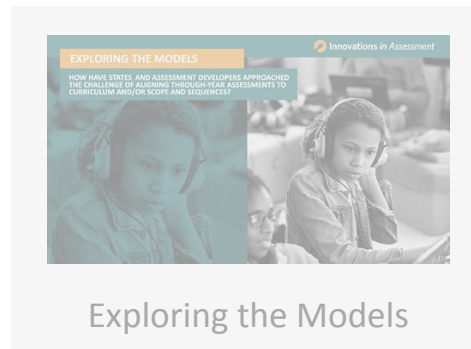
Why Through-Year?



Lessons Learned



Full Report



Exploring the Models



[Return to the landing page](#)